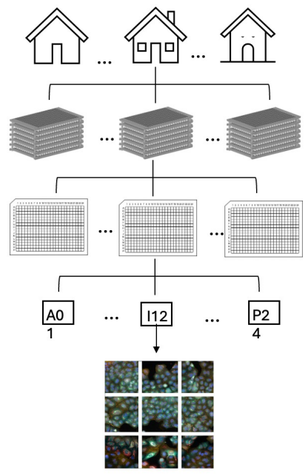
Robust-DINO: Learning Cell Painting Profiles with Technical Artifact Correction

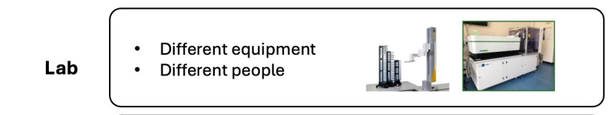
Ghaith Mqawassa and Paula Marin Zapata

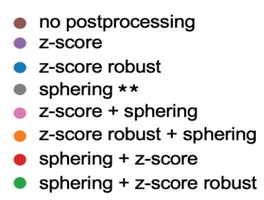
Motivation

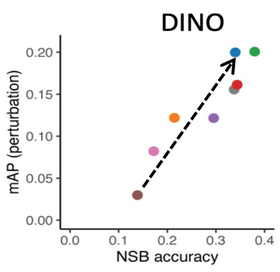
- Cell painting profiles are frequently confounded by technical artifacts (e.g., batch effects or uneven illumination). These variations obscure underlying biological signals, diminishing the assay's accuracy in morphological profiling. Consequently, developing robust analytical pipelines to isolate true biological phenotypes from non-biological noise is critical.

- To alleviate plate-level batch artifacts, previous works applied a background distribution-based transformation. This post-processing technique effectively projected the DINO morphological profiles into a newly established, plate-normalized feature space improving DINO performance in downstream tasks.







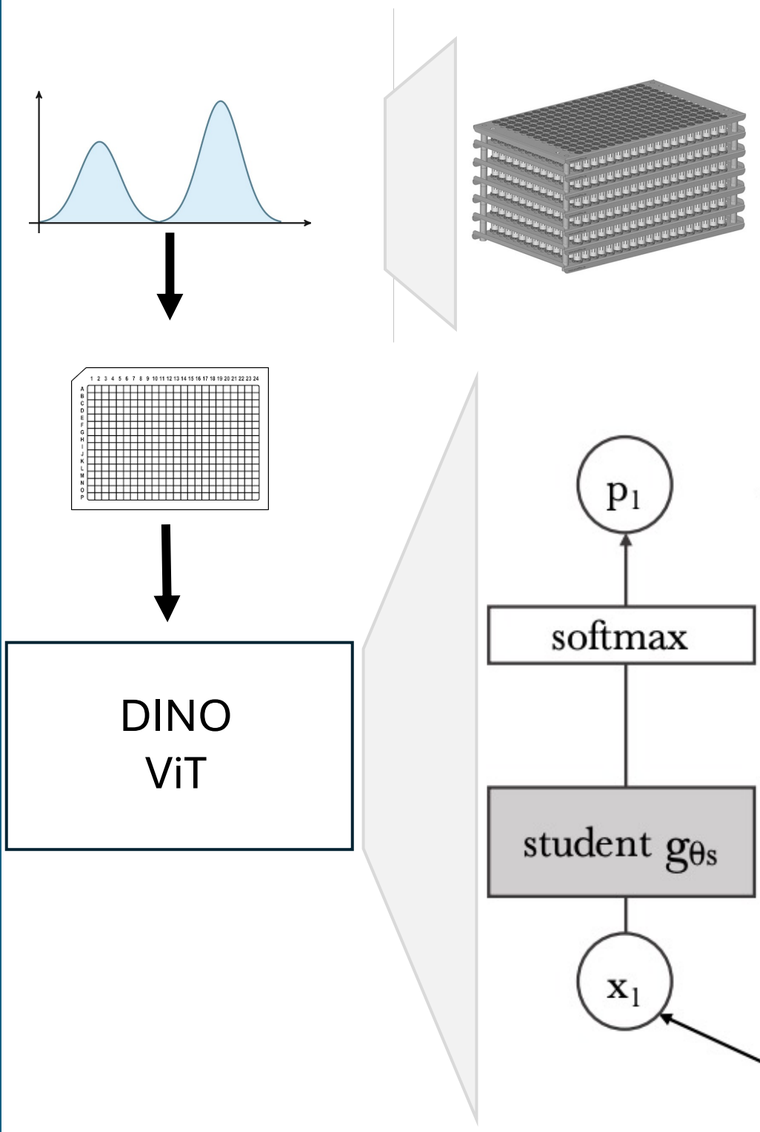


Kim, V., Adaloglou, N., Osterland, M. *et al.* Self-supervision advances morphological profiling by unlocking powerful image representations. *Sci Rep* 15, 4876 (2025). https://doi.org/10.1038/s41598-025-88825-4

Hypothesis

We extract experimental plate statistics during the training phase. Subsequently, we apply a plate-based robust z-score transformation to project the U2OS cell images into a plate-corrected space. By incorporating this normalization into the DINO (Self-Distillation with No Labels) self-supervised training pipeline, we hypothesize that the Vision Transformer (ViT) will prioritize learning intrinsic biological features over exploiting confounding plate-specific shortcut signals for feature discrimination

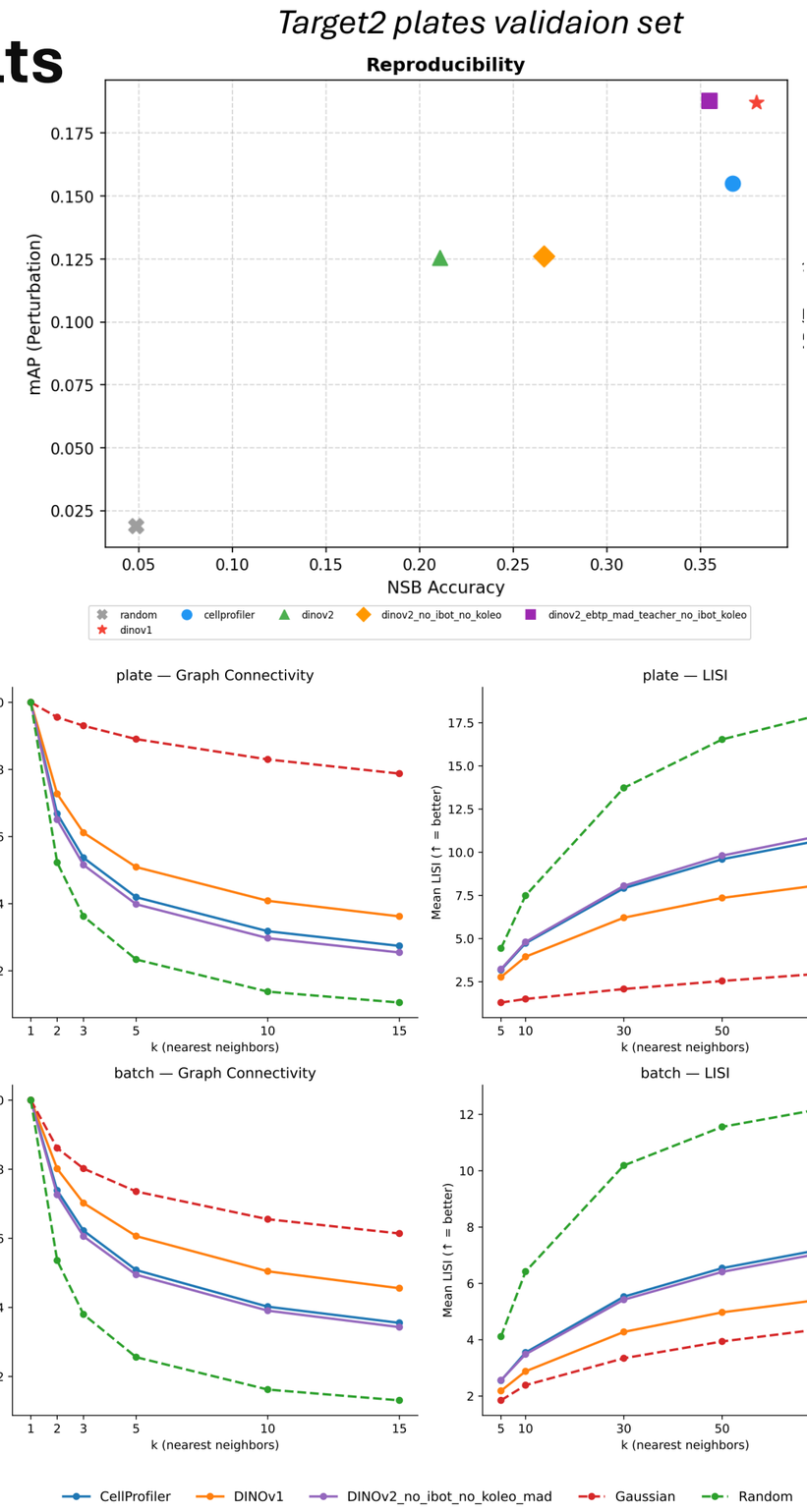
Method



Robust z-scoring

$$f_{\text{norm}} = \frac{\Phi(f_{\text{well}}) - med(\Phi(f_{\text{well}}))}{MAD(\Phi(f_{\text{well}}))}$$

Results



Findings

- ✓ DINOv1 is better than DINOv2 in downstream performance on reproducibility metrics in Target2 plates validation set
- ✓ Robust-DINOv2 performs on par with DINOv1 on reproducibilty metrics and achives **33-40**% increase in reproducibility metrics over DINOv2
- ✓ Robust-DINOv2 is more robust towards batch effects at the plate level compared to other baselines and doesn't exploit counfounding plate-specific shortcut signals for feature discrimination as much as other baselines